转换概率对语音统计学习效应的独立作用

于文勃[1]，王璐[1]，王天琳[2]，梁丹丹[1 3]

[1]南京师范大学；[2]纽约州立大学奥尔巴尼分校；[1 3]中国科学技术大学

摘 要：以往的语音统计学习研究中，目标词的词频和转换概率都显著高于跨界词或非词，学习效应的来源并不明确。为考察转换概率在统计学习中的独立效应，实验一采用 2（词长：两音节，三音节）×3（迫选对类型：目标词和跨界词，目标词和非词，跨界词和非词）的被试内设计，结果发现，被试在三音节跨界词和非词迫选对中的正确率显著高于两音节相应迫选对，在三音节目标词和跨界词迫选对中的正确率边缘显著低于两音节相应迫选对。在实验二中，以音节为单位随机拼接人工语言，发现被试对实验一中所有类型的词语均未产生学习效应。由于实验一中对比情况仅有转换概率有所区别，实验二排除了音节组合偏好的干扰，因此本研究证明了转换概率在统计学习中的独立作用。

关键词：转换概率，词频，统计学习，迫选任务

1 引言

统计学习指个体能够从外界输入的时间信息和空间信息中发现统计信息并以此信息学习新事物的过程（于文勃等，2021a，2021b；Saffran et al., 1996; Frost et al., 2020），统计学习被认为是人类的一项重要认知能力，和口语词切分、词汇语义习得等方面都有较强的联系（Bogaerts et al., 2020; Raviv & Arnon, 2018; Saffran & Kirkham, 2018; Newport, 2016; Siegelman, 2020; Estes et al., 2007，2015）。虽然统计学习机制在不同模态下都有一定的表现，但在语音模态下的研究最为充分（Mirman et al., 2008; Gómez et al., 2017; Wang & Saffran, 2014）。由前向后的转换概率（transitional probability, TP）是统计学习中核心的概率信息之一，其计算公式为：

$$TP_{(X|Y)} = frequencyXY/frequency\,X$$

研究者假设人类大脑可以追踪音节间的转换关系，对高转换概率的音节组合记忆表征更加牢固，达到学习的目的。在编制实验材料过程中，目标词会按照一定的规律进行伪随机拼接，并作为一个整体出现，它们内部相连音节的转换概率较高，如在图 1 中，每个字母代表一个音节，目标词 ABC 内部 AB 和 BC 的转换概率为 1，而两个目标词间前后连接的音节组合被称作跨界词，由于目标词 ABC 后会随机出现其他目标词 JKL、GHI 或 DEF，因此跨界词 CJK 中前两个音节的转换概率为 1/3，后两个音节的转换概率仍旧为 1。在学习阶段后，对被试的学习效果进行测试，如果被试能够区分目标词和跨界词或非词（非词由不同目

标词的音节拼接而成，内部转换概率和词频均为 0），那么就认为出现了学习效应。

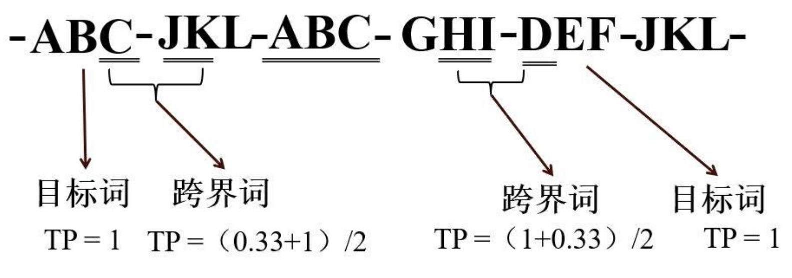


图 1 语音统计学习材料示意图

来自婴幼儿、成人和二语学习者的实验都验证了统计学习机制的假设（Saffran et al., 1996; Estes et al., 2015; Estes & Lew-Williams, 2015; Erickson, Thiessen, & Estes, 2015; Potter, Wang, & Saffran, 2017），但很少提及个体是不是只依靠转换概率进行学习的。如上图所示，虽然目标词的转换概率明显高于跨界词，但是它们在人工语言中出现的频次也更高，假设使用 4 个无意义三音节目标词合成人工语言，目标词间拼接处所形成的跨界词转换概率为 0.33，那么当每个目标词出现 100 次时，跨界词（CJK，HID）在人工语言中出现的次数为 100×0.33=33，也就是说目标词和跨界词在转换概率和词频两个概率信息上都有明显的区别，转换概率和词频的复杂关系导致我们无法给出转换概率是否能够独立影响统计学习效应的证据。

一些研究试图在分离词频因素后考察转换概率对学习效应的影响（Estes, 2012；Aslin et al., 1998）。在婴幼儿实验中，研究者通常是在拼接人工语言过程中将一半的目标词呈现次数减半，随后在转头偏好任务中，仅将高频目标词组合成的跨界词和低频目标词作为测试材料，由于它们在人工语言中的词频相同，因此如果婴儿表现出去习惯化效应，那么就可以证明个体只依靠转换概率就可以完成统计学习任务。此外，Endress 和 Langus（2016）通过对比多个实验的结果发现转换概率在影响统计学习的效应应该大于频次。然而，为数不多的研究仍旧存在一些不足之处。首先，大多数研究在拼接人工语言时都采用相等长度的目标词（两音节或三音节），这一方面并不符合自然语言词长不等的情况（Frost et al., 2020），而且会导致在学习阶段初期被试就产生节奏期待，很容易就发现所有词语都是相等长度的，从而夸大学习效应（Hoch et al., 2013）。其次，由等长度目标词合成的人工语言大大限制了测试词语转换概率的变化范围。以由 4 个三音节目标词合成人工语言为例，所有跨界词的转换概率为 0.67（（1+0.33）/2），目标词转换概率为 1，在迫选任务（转头偏好任务）中只能考察这一种对比情况；然而，如果转换概率确实可以独立影响学习效应，那么在目标词、跨界词和非词两两组成迫选对中，被试的迫选正确率会有所不同，但目前还没有见到这方面的实证研究。

最后，于文勃等（2021b）采用组间设计的方法考察转换概率的独立效应，在实验中通过增加填充词语来降低词语的转换概率，从而保证选项在词频信息上没有区别，而仅在转换概率上有所不同。虽然在成人实验中发现转换概率的提高能够促进统计学习效应，但是这一研究中高低转换概率是组间变量，高转换概率条件下的学习时间较短，低转换概率条件下由于增加了填充词而导致学习时间较长，该条件的目标词在记忆系统中受到了更多的干扰和抑制，所以学习效应较低的解释原因也不唯一。

本研究通过设计混合长度的目标词语来解决以上问题。首先，我们采用两音节和三音节目标词拼接成混合长度的人工语言，在迫选任务中将分别设置两种长度的目标词、跨界词和非词，在相同长度前提下进行两两配对形成三种迫选对类型（目标词和跨界词迫选对，目标词和非词迫选对，跨界词和非词迫选对）。保证每一个迫选试次中两个选项长度相同的原因是普通话母语者存在韵律期待，偏好双音节结构（于文勃等，2021b；裴雨来，2016）。其次，我们重点对比被试在不同长度条件下相应迫选对的正确率差异，这是因为在两音节的目标词和跨界词迫选对中，两个选项在词频和转换概率上都有差异，但是词频的差异在三音节的概率迫选对中仍旧存在，因此当我们再比较同类迫选对在不同长度下的正确率差异时，就可以分离出转换概率的独立效应（六类迫选选项的词频和转换概率详见表 1）。最后，本研究的两个实验都是被试内设计，所有被试均完成相同的学习任务，因此不会因为学习时间长短不同带来额外变量。在实验一中，我们以目标词为单位合成人工语言，虽然目标词都是由无意义音节组成，但这种随机组合方式可能会影响被试的主观感受，实验二以音节为单位采用相同的规则合成人工语言（Toro et al., 2011），迫选任务中使用和实验一完全一致的材料，如果在实验二中，被试没有表现出明显的学习效应，则说明实验一中的效应完全来自于人工语言中目标词所承载的转换概率。

2 实验一：转换概率对统计学习效应的独立作用

2.1 方法

2.1.1 被试

共有 40 名被试参与实验，男性被试 15 名，被试年龄范围 19～25 岁，所有被试母语均为汉语普通话。实验要求被试学习没有接触过的人工语言，为了排除被试二语经验、音乐经验对实验结果的影响，所有被试均非外国语专业和音乐相关专业；最后，为了避免被试猜测实验目的，排除了心理学专业学生。实验前，被试签署知情同意书，实验结束后被试获取少量报酬，本研究经过校伦理委员会审查（×××2022060023）。

2.1.2 实验设计

采用学习-测试范式考察被试的语音统计学习能力，实验设计为2（词长：两音节，三音节）×3（迫选对类型：目标词和跨界词迫选对，目标词和非词迫选对，跨界词和非词迫选对）的被试内设计。在具体的分析中，首先检验被试在迫选对上的正确率与随机水平（0.5）的差异性，随后进行方差分析检验转换概率的独立效应。

2.1.3 材料和实验程序

在学习阶段，由10个音节构造目标词来合成人工语言，所有音节均来自于文勃等（2021b）的研究，由一名女性普通话母语者录制，经过标准化后音节时长 300ms，强度 70dB，基频 266Hz。参考以往研究设计，这些音节在普通话第一声声调下无意义，随后将 10 个音节随机组成 4 个目标词（Gómez et al., 2017；于文勃等，2021b），音节和音节之间没有任何时间间隔。根据以往统计学习实验的要求（Saffran et al., 1996），按照伪随机的方式拼接人工语言，其中每个目标词后不能立刻出现该目标词，并且出现其他目标词的概率相同（本研究中为1/3），在人工语言中每个目标词出现 100 次，保证人工语言前后两段目标词均匀分布，人工语言时长 5 分钟。随后，分别合成跨界词和非词各 4 个，前者要求目标词包括前后相连两个目标词的部分音节，具有一定的转换概率，后者要求词内音节不会相连出现在人工语言中，因此转换概率为 0。编制迫选对过程中，为了保证每一个迫选试次内的选项仅在转换概率上存在差异，因此每个迫选对均是由相同长度的选项构成，同时还要将两个选项调换顺序呈现两次，以排除选项顺序对实验结果的影响。迫选对中的六类词的转换概率、在人工语言中的频次如表 1。

表 1 实验中六类选项的转换概率和频次

词类	词	词频	转换概率
两音节目标词	remei，rouse	100	1
两音节跨界词	meirou，sere	33	0.33
两音节非词	refo，rouruo	0	0
三音节目标词	nueruote，diafolai	100	1
三音节跨界词	tediafo，lainueruo	33	0.67
三音节非词	nuemeilai，diasete	0	0

实验程序由 E-prime 呈现，被试佩戴耳机完成，电脑音量由被试自主调节，在 30%～40%范围内。实验程序包括练习实验和正式实验，练习实验前由主试讲解实验要求和指导语，并播放 5s 中的人工语言，随后呈现 3 个迫选试次帮助被试了解实验流程，练习实验中的材料在正式实验中不会出现。正式实验被试需要学习人工语言 5 分钟，随后完成 24 个迫选试次，

每个迫选试次中，两个选项仍旧采用音频的方式播放，中间间隔 500ms，全部播放结束后，屏幕呈现提示语，要求被试从听到的两个词语中选择更为熟悉的那个，按“1”或“2”键进行选择。实验全程大约需要 10 分钟，流程示意图如图 2。

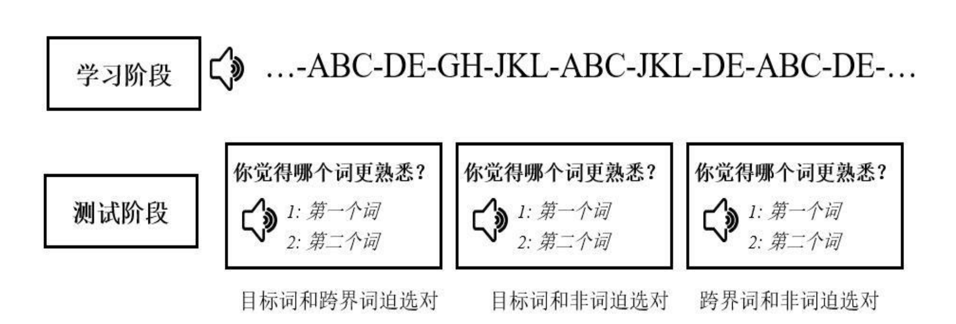


图 2 实验流程示意图

2.2 结果

2.2.1 混合词长条件下的统计学习效应

在统计学习研究中，首先要通过迫选正确率和随机水平的差异比较来判断被试是否实现了统计学习，跨界词和非词都不是被试所要学习的人工词语，在本研究中的前两种迫选对类型中它们都是错误答案；不过，在跨界词和非词迫选对中，由于跨界词的转换概率更高且在人工语言中出现了一定的次数，因此如果相比于非词被试能够更多地选择跨界词，则也说明被试进行了统计学习。

单样本 T 检验结果表明，被试在目标词和跨界词迫选对中，迫选正确率边缘显著高于随机水平，t（39）= 1.97，$p = 0.056$，$M = 0.57$，差异的 95%CI:：[-0.00，0.14]，$d = 0.31$，在目标词和非词迫选对中，迫选正确率显著高于随机水平，t（39）= 4.53，$p < 0.001$，$M = 0.63$，差异的 95%CI:：[0.07，0.19]，$d = 0.72$，在跨界词和非词迫选对中，迫选正确率边缘显著高于随机水平，t（39）= 1.86，$p = 0.070$，$M = 0.55$，差异的 95%CI:：[-0.00，0.10]，$d = 0.23$。被试的总体正确率显著高于随机水平，t（39）= 3.71，$p < 0.001$，$M = 0.55$，差异的 95%CI:：[0.04,0.13]，$d = 0.59$。以上结果表明，被试在学习过程中已经能够对有转换概率的结构（目标词和跨界词）进行记忆表征，表现出统计学习效应，如图 3。

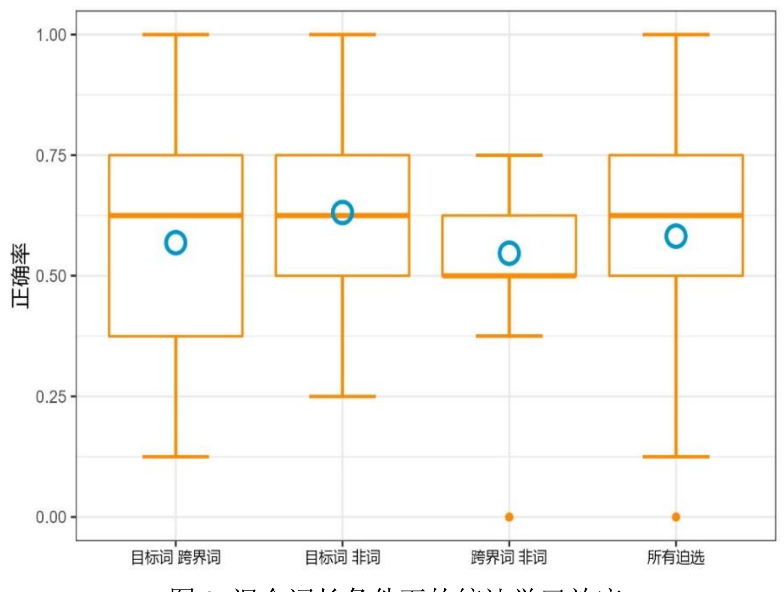


图 3 混合词长条件下的统计学习效应

2.2.2 转换概率对学习效应的影响

两因素被试内方差分析显示，追选对类型主效应边缘显著（F（1.75，68.13）= 3.15，p = 0.056，η^2_g = 0.02），词长主效应不显著（F（1，39）= 1.39，p = 0.245），追选对类型和词长交互作用显著（F（1.83，71.31）= 4.80，p = 0.013，η^2_g = 0.03）。对追选对词类的主效应事后检验表明，被试在目标词和非词的追选对上的正确率显著高于跨界词和非词（t(39) = 2.68，p = 0.032）。两个因素交互作用的简单效应分析表明，在目标词和跨界词追选对中，两音节条件下的正确率边缘显著高于三音节条件（t(39) = 1.71，p = 0.095），但在跨界词和非词条件下，两音节条件下的正确率显著低于三音节条件（t(39) = -2.60，p = 0.013），在目标词和非词追选对中，词长的简单效应不显著（t(39) = -1.35，p = 0.185）。此外，在两音节词长条件下，目标词和跨界词追选对（t(39) = 2.34，p = 0.093）以及目标词和非词追选对（t(39) = 2.52，p = 0.048）的正确率都（边缘）显著高于跨界词和非词追选对。在三音节条件下，仅有目标词和跨界词追选对的正确率显著低于目标词和非词追选对（t(39) = -2.72，p = 0.029），其他条件差异未达到显著水平，以上分析中，涉及到三个及三个水平以上的检验时，均采用了 bonferroni 矫正。被试在六种追选对下的正确率和标准差如表 2 和图 4。

表 2 三类迫选对的正确率和标准差

		目标词和跨界词		目标词和非词		跨界词和非词	
		两音节	三音节	两音节	三音节	两音节	三音节
实验一	M	0.61	0.53	0.60	0.66	0.49	0.61
	SD	0.27	0.28	0.23	0.24	0.20	0.23
实验二	M	0.55	0.49	0.41	0.52	0.49	0.43
	SD	0.26	0.22	0.27	0.22	0.23	0.24

值得注意的是，被试在三音节目标词和跨界词迫选对和两音节跨界词和非词迫选对上的正确率均与随机水平差异不显著，t（39）= 0.57，M = 0.53，p = 0.570，t（39）= -0.40，M = 0.49，p = 0.689，这两个结果说明被试在这些类迫选对上没有表现出学习效应。

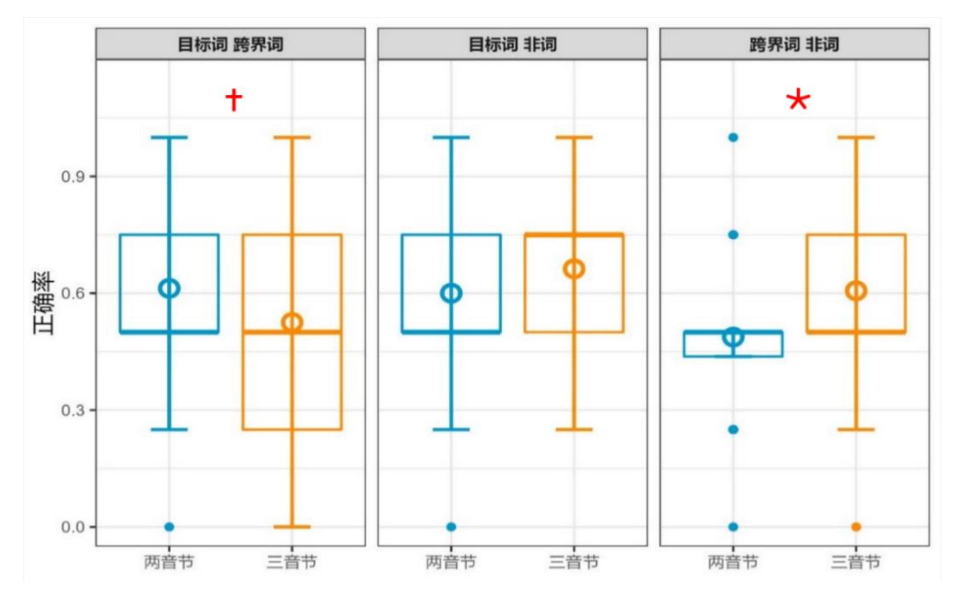


图 4 转换概率对学习效应的影响

2.3 实验一讨论

在目标词和跨界词以及跨界词和非词组成的迫选对中，被试的迫选正确率均在 0.55 上下，和随机水平差异边缘显著；但被试在目标词和非词的迫选对中，被试的迫选正确率达到 0.63，显著高于随机水平。在重复测量方差分析中，我们发现被试在三类迫选对正确率差异边缘显著，尤其是目标词和非词的迫选对中正确率显著高于跨界词和非词的迫选对。分析其原因，目标词、跨界词和非词三类词在学习材料中出现的频次和转换概率逐渐下降，从统计学习的记忆模型来看（Thiessen & Erik, 2017; Lukics & Lukács, 2021; Isbilen et al., 2020; 2022），高转换概率的结构也就是目标词，在人工语言中出现频次更多，被试有更多的机会对它们进行表征；而跨界词在人工语言中出现频次较低，转换概率也相对较低，被试对它们的记忆表

征相对较弱；最后，非词并未出现在人工语言中，因此被试对它们的熟悉度或再认可能性都最低，这些原因共同导致了被试在不同类型迫选对的正确率存在一定的差异。

相比于以往研究，实验一在实验材料上进行了改进，两音节跨界词和三音节跨界词的转换概率并不相同（两音节：0.33，三音节：0.67），但它们在人工语言中出现的频次相同。当它们分别和长度相同的目标词进行迫选时，虽然在转换概率和词频上都有差别，但是当我们再比较两种长度迫选试次的正确率时，则只在转换概率上存在差异。相比于三音节跨界词，两音节跨界词的转换概率较低，它们对两音节目标词造成的干扰较小，被试可以轻松地选择出两音节目标词，因此两音节目标词和跨界词迫选的正确率边缘显著高于三音节目标词和跨界词的迫选对。同理，两音节跨界词转换概率较低，三音节跨界词转换概率较高，因此后者在和非词（转换概率为 0，词频也为 0）进行配对时，被试较容易进行选择出正确答案，最终导致三音节跨界词和非词迫选正确率显著高于两音节条件。因此，实验一交互作用结果进一步说明当词频差异固定时，转换概率能够影响学习效应。

在实验二中，我们以音节为单位合成人工语言，采用相同的迫选材料进行实验。如果被试是因为对目标词或跨界词有偏好，那么他们仍旧会表现出明显的“学习效应”，相反，他们选择正确答案的可能应不超过随机水平（0.5）。

3 实验二：随机人工语言的学习效应检验

3.1 方法

3.1.1 被试

招募被试的标准同实验一，共有 38 名被试参与实验二，其中男性被试 10 名，被试年龄范围为 19～26 岁。由于实验程序问题，一名被试的数据被覆盖，最终有 37 名被试的行为数据进入分析。

3.1.2 实验设计

仍旧采用学习-测试范式考察被试的语音统计学习能力，实验自变量和因变量同实验一保持一致。由于人工语言以音节为单位合成，所有目标词和跨界词在学习阶段中的转换概率均较低，不应该导致学习效应。迫选任务中以实验一的答案作为标准检验被试在随机人工语言条件下是否对之前的正确答案有先验的语言偏好。

3.1.3 材料和实验程序

实验二中学习阶段的人工语言以实验一中的 10 个无意义音节为单位合成，每个音节后不能重复出现该音节，可以是其他 9 个音节中的任意一个，最终，人工语言由 1000 个音节

拼接而成。实验二的迫选试次和实验一完全一致，经过计算所有选项在人工语言出现频次不超过 20 次，所占比例非常低。实验程序和实验一保持一致，由 E-prime 呈现，包括练习实验和正式实验，时间约 10 分钟。

3.2 结果

单样本 T 检验结果显示，被试在所有试次上的迫选正确率未显著高于随机水平，t（36）= -1.59，M = 0.48，p = 0.121，；此外，被试在目标词和跨界词迫选对中，迫选正确率未显著高于随机水平，t（36）= 0.61，M = 0.53，p = 0.549，在目标词和非词迫选对中，迫选正确率未显著高于随机水平，t（36）= -1.23，M = 0.46，p = 0.226，在跨界词和非词迫选对中，迫选正确率未显著高于随机水平，t（36）= -1.54，M = 0.53，p = 0.133。随后，结合词长变量检验被试在六种条件下是否表现出明显的学习效应，结果发现被试的迫选正确率均未显著高于随机水平（p.s. > 0.05），正确率和标准差如表 2。

3.3 实验二讨论

实验二以音节为单位合成人工语言作为学习材料，考察缺失转换概率因素时被试是否对实验一中的“答案”仍旧表现出更高的熟悉性，结果发现无论在总体还是六种水平下被试的迫选正确率均未显著高于随机水平，这说明实验一中的学习效应不是来自于被试对目标词、跨界词的先验语言经验，而是来自于学习阶段对转换概率信息的习得。

4 总讨论

以往的统计学习研究中，注重检验个体能否依靠统计信息实现统计学习或词切分，较少关注个体对哪些统计信息进行了利用。本研究的目的在于考察转换概率高低是否会影响被试的学习效应，从而验证其对统计学习效应的独立贡献。实验一中，我们设置了不同长度的目标词、跨界词和非词，结果发现被试能够较好地完成迫选任务，表现出统计学习效应；同时，不同类型迫选对中选项间转换概率的差异确实影响被试的迫选正确率。实验二中，我们通过以音节为单位编制的人工语言发现被试在实验一中所表现出的学习效应并非来自对无意义目标词和跨界词的先验偏好。结合两个实验结果，说明转换概率能够独立于词频影响语音统计学习任务的结果。

4.1 统计学习效应

经典的统计学习任务为了验证个体能否仅依靠音节间的转换概率来实现语流切分，通常只选用两音节词（Mirman et al., 2008; Gómez et al., 2017；Estes et al., 2015）或三音节词（Saffran et al., 1996; Wang & Saffran, 2014）来合成人工语言。但近年来，一些学者提出，为

了更好地贴近自然语言中词长不等的情况（Frost et al., 2020; Saffran & Kirkham, 2018），同时避免韵律期待效应（Hoch et al., 2013），应该使用长度不等的词语合成实验材料（于文勃等，2021b）。在实验一中我们发现被试在所有迫选试次下的正确率显著高于随机水平，达到了0.58，仍旧表现出了学习效应，与以往研究结果一致（Palmer & Mattys, 2016; Gómez et al., 2017；Estes et al., 2015; 于文勃等，2021b），说明即使在较为复杂的语言环境下，个体仍旧可以通过追踪语言内部的概率信息实现词语切分。实验一的设计更进一步地验证记忆模型在解释统计学习效应的合理性。如表 1 和图 4 所示，我们采用了六类词语组成迫选对，结果发现被试在转换概率和词频差异较大的迫选对中正确率较高(如三音节目标词和非词迫选对)，而在差异较小的迫选对中正确率较低，甚至和随机水平差异不显著(如两音节跨界词和两音节非词)。相比于以往研究，本研究为统计学习的记忆模型理论提供了更精细的实验证据。

4.2 转换概率的独立效应

统计学习的理论源来自于 Harris（1954，1955）对自然口语中语素分布规律的归纳，作者认为个体可以通过发现程序习得语言或切分语流，而发现程序的关键就是对语流中的统计信息进行加工。虽然词频和转换概率都可以被看成概率信息，但是它们二者有着本质的区别。对于高转换概率的音节组合，前一个音节可以预测后一个音节的出现，但对于高频次的音节组合来说，它们的转换概率可能并不高，即前一个音节也可能和其他音节组成词，并且多次出现在语言环境中，所以转换概率更能够作为判断语流中词边界的因素。本研究中，虽然在目标词和跨界词的迫选中，两音节条件只是边缘显著高于三音节条件，而且效应量也未达到中等程度，但单样本 T 检验的结果显示被试的三音节迫选正确率（0.53）与随机水平差异不显著，这说明在这种条件下已经没有了学习效应，被试难以分辨三音节目标词和三音节跨界词的区别；相应的，在两音节迫选任务中学习效应仍旧稳定（正确率达到 0.61），说明在这种迫选类型下被试能够很好地区分目标词和跨界词。因此从被试的学习效果来看可以认为在词频被控制后，音节间的转换概率仍旧能够影响语音统计学习效应。

Endress 和 Mehler（2009）在人工语言中设置了“幽灵词”，这些词虽然没有以音节组合的形式出现过，但是音节间的转换概率和目标词相同，实验结果显示被试对幽灵词和目标词的迫选正确率与随机水平差异不显著，但被试对幽灵词的喜好程度高于跨界词。这一结果说明被试首先是以转换概率作为词语切分和存储的标准，随后才是频次等信息（Perruchet & Poulin-Charronnat, 2012）。统计学习中转换概率的独立作用在自然语言中也可以找到证据。口语语流中缺少明确的词间边界，因此音节间的统计信息对于词切分尤为关键。在英语口语

语料库[1]中“there is”和“there are”在每百万词中出现的频次高达 458.6 和 433.82，二者都是高频结构；但“there”后出现“is”的可能并不高[2]，即音节间的转换概率不高。虽然“there is”具备很高的频次，但由于转换概率不高，所以更可能被切分成两个单词而非一个单词的两个词内音节。

4.3 研究启示

本研究结果有助于思考语言使用视角下的汉语词边界问题。从语言学角度看，词的定义是能够独立运用的最小音义结合体，它不仅是个体在头脑中存储的基本单位，还是在构造句子过程中不可再拆分的结构。在英语中，语言学意义上的词和个体头脑中的词概念一一对应，如 a，apply，red 既是从语法角度划分的词也是语言使用角度最直接感受到的词结构。汉语普通话中，在语法的界定中，“打球”、“心软”的结构是短语或词组，但对普通话母语者来说这些结构在心理上是词[3]，例如 Cai 等人（2010）通过建立电视连续剧字幕语料库统计汉语口语词频信息，就包括了“打球”、“吃饭”这些词组结构的词频数据。在音节“打”后还可能出现音节“劫”、“听”、“中”、“入”、“电话”等连谓、动补和动宾等结构，事实上，“打球”结构中两个音节间的转换概率并不高，既然本研究发现口语加工中前一个音节对后一个音节的预测性（转换概率）能够影响大脑对词边界的识别，那么在类似结构中个体很可能只将“打”切分出来作为一个结构单位加工[4]，而不是将“打球”作为一个整体进行加工。本研究的结果能够为汉语词边界划分在语言学理论和语言使用视角下的不统一问题提供参考。

5 结论

本研究发现，当词频固定时，转换概率仍旧能够影响统计学习的效应，被试仍旧可以识别出高转换概率的目标词或跨界词，说明转换概率对统计学习的影响在一定程度上独立于词频。

---

[1] 数据来源：British National Corpus Online service，http://bncweb.lancs.ac.uk

[2] 在口语中，their 和 there 的发音相同，因此/ðɛə/后出现的音节种类更多，进一步降低了 there is 的转换概率。

[3] 在相关研究中，有学者称之为主观词（闫国利等，2013）、离合词（朱佳蕾，刘凤樨，2020；庞加光，张韧，2022）或韵律词（裴雨来，2016）。

[4] Cai（2010）等人的文章中包括了“打”作为一个词的词频信息。

参考文献

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.

Bogaerts, L., Frost, R., & Christiansen, M. H. (2020). Integrating statistical learning into cognitive science. *Journal of Memory and Language,115*,104167.

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*(3), 351−367.

Endress, A. D., & Langus, A. (2016). Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive Psychology, 92*, 37.

Erickson, L. C., Thiessen, E. D., & Estes, K. G. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language*, 72, 49-58.

Estes, K. G. (2012). Infants generalize representations of statistically segmented words. *Frontiers in Psychology*, 3, 447.

Estes, K. G., & Lew-Williams, C. (2015). Listening through Voices: Infant Statistical Word Segmentation across Multiple Speakers. *Developmental Psychology, 51*(11), 1517–1528.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can Infants Map Meaning to Newly Segmented Words? Statistical Segmentation and Word Learning. *Psychological Science, 18*(3), 254–260.

Estes, K. G., Gluck, C. W., & Bastos, C. (2015). Flexibility in statistical word segmentation: finding words in foreign speech. *Language Learning and Development, 11*(3), 252-269.

Frost, R., Armstrong. B. C., & Christiansen. M. H. (2020). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 145*(12), 1128-1153.

Gómez, D. M., Mok, P., Ordin, M., Mehler, J., & Nespor, M. (2017). Statistical speech segmentation in tone languages: The role of lexical tones. *Language & Speech, 61*(1), 84–96.

Harris, Z. S. (1954). Distributional structure. *Word, 10*(2-3), 146-162.

Harris, Z. S. (1955). From phoneme to morpheme. *Language, 31*(2), 190-222.

Hoch, L., Tyler, M. D., & Tillmann, B. (2013). Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychonomic Bulletin & Review, 20*(1), 142–147.

Isbilen, E. S., Mccauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: a memory-based approach to statistical learning. *Cognitive Science*, 44(7).

Isbilen, E.S., Frost, R., Monaghan, P., & Christiansen, M. H. (2022). Statistically-based chunking of non-adjacent

dependencies. PsyArXir.

Lukics, K. S., & Lukács, G. (2021) Tracking statistical learning online: Word segmentation in a target detection task. *Acta Psychologica, 215*(7):103271.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1), 271-280.

Newport, E. L. (2016). Statistical language learning: Computational, maturational, and linguistic constraints. *Language and Cognition, 8*(3), 447−461.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology, 69*(12), 2390-2401.

Pang, J., & Zhang, R. (2022). Chinese Splittable Compounds from the Perspective of Baseline/Elaboration Model. *Journal of Foreign Languages, 45*(2),72

[庞加光，张韧.（2022）.基线/阐释模型下的汉语离合词现象研究.*外国语(上海外国语大学学报)*(02),72-83.]

Pei, Y. (2016). *Prosodic words in Chinese*. Beijing：Beijing Language and Culture University Press.

[裴雨来. (2016). *汉语的韵律词.* 北京：北京语言大学出版社.]

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language*, *66*(4), 807−818.

Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive science*, 41(S4), 913-927.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: modality‐based differences in the effect of age. *Developmental Science*, 21(4), e12593.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, 69(1), 181–203.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926−1928.

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass, (14)*, e12365.

Thiessen, E. D., & Erik, D. (2017). What's statistical about learning? insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372.*

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The Contribution of Language-Specific Knowledge in the Selection of Statistically-Coherent Word Candidates. *Journal of Memory and Language*, *64*(2), 171–180.

Wang, T. L., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology, 5*.

Yan, G., Zhang, L., Sun, S., & Bai, X. (2013). The Representation and Processing of Subjective Words for Chinese Readers. *Acta Psychologica Sinica*, *45*(4), 379.

[闫国利，张兰兰，孙莎莎，白学军.(2013).汉语“主观词”的表征及其加工. *心理学报*(04),379-390.]

Yu, W., Wang, L., Cheng, X., Wang, T., Zhang, J., & Liang, D. (2021a). The influence of linguistic experience on statistical word segmentation. *Advances in Psychological Science, 29*(5), 787.

[于文勃，王璐，程幸悦，王天琳，张晶晶，梁丹丹.（2021a）.语言经验对概率词切分的影响.*心理科学进展，* 29(05)：787-795.]

Yu, W., Wang, L., Qu, X., Wang, T., Zhang, J., & Liang, D. (2021b). Transitional probabilities and expectation for word length impact verbal statistical learning. *Acta Psychologica Sinica, 53*(6), 565.

[于文勃，王璐，瞿邢芳，王天琳，张晶晶，梁丹丹.（2021b）. 转换概率和词长期待对语音统计学习的影响. *心理学报，53*(06)：565-574.]

Zhu, J., & Liu, F. (2020) Separable Verb－Object Compounds in Chinese and the Division of Labor between Morphology and Syntax: The Analysis of Cognate Object Ｒevisited. *Contemporary Linguistics*，*Contemporary Linguistics*, *22*(3), 317-334.

[朱佳蕾，刘凤樨.(2020).离合词与词法句法的分工——再议“同源宾语说”. *当代语言学*(03),317-334.]

The independent effect of transitional probability on verbal statistical learning

Abstract: In a typical SL task, participants are first exposed to a nonsensical artificial language for 5~10 mins and then asked to finish a 2 alternative forced choice task (2AFC). Transitional probabilities (TPs), the core concept in SL, represent the predictably between syllables. In a given artificial language, syllables within a target word occur together more frequently, resulting in higher TPs than those of the syllables that span across word boundaries. The latter is referred to as partwords and consist of lower TPs. After the exposure phase, participants enter the test phase and are presented with a partword and target word in each trial of the 2AFC task. If the accuracy across participants is higher than chance level, it is assumed that learning has occurred. However, studies have also shown that factors other than TPs, such as word token frequency, word length variations (or the lack thereof) also impact SL performance in such tasks. To date, these factors as well as their interactive effects remain under studied.

In the experiment one, we aimed to investigate whether TPs affect SL learning performance when controlling for target words' and partwords' token frequencies. In doing so, we created the artificial language by randomizing the order of two trisyllabic words and two disyllabic words. During the 2AFC task, three types of items (target word, partword, and nonword) were paired together, with two items in equal length in each trial. There were 24 trials in the test. 40 native Mandarin monolinguals participated in the experiment; they first listened to the artificial language for 5mins and then finished the 2AFC task. In the experiment two, an artificial language was generated with 10 syllables and presented in exposure phase, to examine whether experiment one's learning effect came from the TP or participants' prior language bias.

Results in experiment one showed that the accuracy of all trials was significantly higher than chance (0.5) at the group level, suggesting that participants were able to segment the artificial language of mixed word lengths. Participants were also marginally better at choosing target words over partwords, and partwords over nonwords. To investigate the independent effect of TP in SL, we subset the data by word length and found that participants' accuracy choosing trisyllabic target words over partwords was marginally lower than their choosing disyllabic target words over partword, which suggests that disyllabic words confer advantage in SL for this group of participants. In addition, participants' accuracy in choosing trisyllabic partwords over nonwords was significantly higher than that of disyllabic target partwords over nonwords. In the experiment two, there were no significant learning effect in any levels when the statistical information was absent.

A series of results across two behavior experiments highlight the unique contribution from TPs alone, since accuracy was assessed by controlling for word token frequency and word length. Thus,

the present study suggests that TP exerts effect on verbal SL performance independent of word token frequency. Further studies should take into account more types of statistical rule such as mutual information and backward TP.
Key words: transitional probability; tokens; statistical learning; 2-alternative forced choice task